GENERATION OF PLANTS WITH IMPROVED PATHOGEN RESISTANCE

BACKGROUND OF THE INVENTION

The control of infection by plant pathogens, which can inhibit production of fruits, seeds, foliage and flowers and cause reductions in the quality and quantity of the harvested crops, is of significant economic importance. Pathogens annually cause billions of dollars in damage to crops worldwide (Baker et al. 1997, Science 276:726-733). Consequently, an increasing amount of research has been dedicated to developing novel methods for controlling plant diseases. Such studies have centered on the plant's innate ability to resist pathogen invasion in an effort to buttress the plant's own defenses to counter pathogen attacks (Staskawicz et al. 1995, Science 268:661-667; Baker et al. *supra*).

Although most crops are treated with agricultural anti-fungal, anti-bacterial agents and/or pesticidal agents, damage from pathogenic infection still results in revenue losses to the agricultural industry on a regular basis. Furthermore, many of the agents used to control such infection or infestation cause adverse side effects to the plant and/or to the environment. Plants with enhanced resistance to infection by pathogens would decrease or eliminate the need for application of chemical anti-fungal, anti-bacterial and/or pesticidal agents.

There has been significant interest in developing transgenic plants that show increased resistance to a broad range of pathogens (Stuiver and Custers, 2001, Nature 411:865-8; Melchers and Stuiver, 2000, Curr Opin Plant Biol 3:147-52; Rommens and Kishore, 2000, Curr Opin Biotechnol 11:120-5; Mourgues et al. 1998, Trends Biotechnol 16:203-10). The interaction between *Arabidopsis* and the oomycete *Peronospora parasitica* (downy mildew) provides an attractive model system to identify molecular components of the host that are required for recognition of the fungal parasite (Parker et al.1996 Plant Cell8:2033-46). A number of genes whose mis-expression is associated with altered resistance to *P. parasitica*, as well as other pathogens, have been identified in *Arabidopsis*. Overexpression of the NPR1 gene confers resistance to infection by *P. parasitica* as well as the bacterial pathogen *Pseudomonas syringae* (Cao et al, 1998 Proc Natl Acad Sci U S A 95:6531-6536). CPR6 is semi-dominant mutation implicated in multiple defense pathways (Clarke et al. 1998, Plant Cell 10:557-569). Lsd6 and Lsd7 are dominant mutations that confer heightened disease and result in the development of spontaneous necrotic lesions and elevated levels of salicylic acid (Weymann et al 1995 Plant Cell 7:2013-2022). A number of recessive mutations confer *P. parasitica* resistance,

including ssi2, in the SSI2 gene encoding a stearoyl-ACP desaturase (Kachroo et al. 2001 Proc Natl Acad Sci U S A 98:9448-9453), mpk4, in a MAP kinase gene (Petersen et al. 2000, Cell 103:1111-20), and pmr4 (Vogel and Somerville 2000 Proc Natl Acad Sci U S A 97:1897-1902). The recessive mutations cpr5 and cpr1 also confer resistance to *P. syringae* and cause a dwarf phenotype (Bowling et al 1997 Plant Cell 9:1573-1584; Bowling et al, 1994 Plant Cell 6:1845-1857).

Activation tagging in plants refers to a method of generating random mutations by insertion of a heterologous nucleic acid construct comprising regulatory sequences (e.g., an enhancer) into a plant genome. The regulatory sequences can act to enhance transcription of one or more native plant genes; accordingly, activation tagging is a fruitful method for generating gain-of-function, generally dominant mutants (see, *e.g.*, Hayashi *et al.*, Science (1992) 258: 1350-1353; Weigel *et al.*, Plant Physiology (2000) 122:1003-1013). The inserted construct provides a molecular tag for rapid identification of the native plant whose mis-expression causes the mutant phenotype. Activation tagging may also cause loss-of-function phenotypes. The insertion may result in disruption of a native plant gene, in which case the phenotype is generally recessive.

Activation tagging has been used in various species, including tobacco and *Arabidopsis*, to identify many different kinds of mutant phenotypes and the genes associated with these phenotypes (Wilson *et al.*, Plant Cell (1996) 8:659-671, Schaffer *et al.*, Cell (1998) 93: 1219-1229; Fridborg *et al.*, Plant Cell (1999)11: 1019-1032; Kardailsky *et al.*, Science (1999) 286:1962-1965); Christensen S *et al.*, 9th International Conference on *Arabidopsis* Research. Univ. of Wisconsin-Madison, June 24-28, 1998. Abstract 165). In one example, activation tagging was used to identify mutants with altered disease resistance (Weigel *et al., supra*).

SUMMARY OF THE INVENTION

The invention provides a transgenic plant comprising a plant transformation vector comprising a nucleotide sequence that encodes or is complementary to a sequence that encodes a PPR1 polypeptide or an ortholog thereof. The transgenic plant is characterized by having increased resistance to pathogens controlled by the salicylic acid-dependent resistance pathway relative to control plants.

The present invention further provides a method of producing an altered pathogen resistance phenotype in a plant. The method comprises introducing into plant progenitor cells a vector comprising a nucleotide sequence that encodes or is complementary to a

sequence encoding a PPR1 polypeptide or an ortholog thereof and growing a transgenic plant that expresses the nucleotide sequence. In one embodiment, the PPR1 polypeptide has at least 50% sequence identity to the amino acid sequence presented in SEQ ID NO:2 and comprises an AP2 domain. In other embodiments, the PPR1 polypeptide has at least 80% or 90% sequence identity to or has the amino acid sequence presented in SEQ ID NO:2.

The invention further provides plants and plant parts obtained by the methods described herein.

DETAILED DESCRIPTION OF THE INVENTION

Definitions

Unless otherwise indicated, all technical and scientific terms used herein have the same meaning as they would to one skilled in the art of the present invention. Practitioners are particularly directed to Sambrook *et al.,* Molecular Cloning: A Laboratory Manual (Second Edition), Cold Spring Harbor Press, Plainview, N.Y.,1989, and Ausubel FM *et al.,* Current Protocols in Molecular Biology, John Wiley & Sons, New York, N.Y., 1993, for definitions and terms of the art. It is to be understood that this invention is not limited to the particular methodology, protocols, and reagents described, as these may vary.

As used herein, the term "vector" refers to a nucleic acid construct designed for transfer between different host cells. An "expression vector" refers to a vector that has the ability to incorporate and express heterologous DNA fragments in a foreign cell. Many prokaryotic and eukaryotic expression vectors are commercially available. Selection of appropriate expression vectors is within the knowledge of those having skill in the art.

A "heterologous" nucleic acid construct or sequence has a portion of the sequence that is not native to the plant cell in which it is expressed. Heterologous, with respect to a control sequence refers to a control sequence (*i.e.* promoter or enhancer) that does not function in nature to regulate the same gene the expression of which it is currently regulating. Generally, heterologous nucleic acid sequences are not endogenous to the cell or part of the genome in which they are present, and have been added to the cell, by infection, transfection, microinjection, electroporation, or the like. A "heterologous" nucleic acid construct may contain a control sequence/DNA coding sequence combination that is the same as, or different from a control sequence/DNA coding sequence combination found in the native plant.

As used herein, the term "gene" means the segment of DNA involved in producing a polypeptide chain, which may or may not include regions preceding and following the coding region, *e.g.* 5' untranslated (5' UTR) or "leader" sequences and 3' UTR or "trailer" sequences, as well as intervening sequences (introns) between individual coding segments (exons) and non-transcribed regulatory sequence.

As used herein, "recombinant" includes reference to a cell or vector, that has been modified by the introduction of a heterologous nucleic acid sequence or that the cell is derived from a cell so modified. Thus, for example, recombinant cells express genes that are not found in identical form within the native (non-recombinant) form of the cell or express native genes that are otherwise abnormally expressed, under expressed or not expressed at all as a result of deliberate human intervention.

As used herein, the term "gene expression" refers to the process by which a polypeptide is produced based on the nucleic acid sequence of a gene. The process includes both transcription and translation; accordingly, "expression" may refer to either a polynucleotide or polypeptide sequence, or both. Sometimes, expression of a polynucleotide sequence will not lead to protein translation. "Over-expression" refers to increased expression of a polynucleotide and/or polypeptide sequence relative to its expression in a wild-type (or other reference [e.g., non-transgenic]) plant and may relate to a naturally-occurring or non-naturally occurring sequence. "Ectopic expression" refers to expression at a time, place, and/or increased level that does not naturally occur in the non-altered or wild-type plant. "Under-expression" refers to decreased expression of a polynucleotide and/or polypeptide sequence, generally of an endogenous gene, relative to its expression in a wild-type plant. The terms "mis-expression" and "altered expression" encompass over-expression, under-expression, and ectopic expression.

The term "introduced" in the context of inserting a nucleic acid sequence into a cell, means "transfection", or "transformation" or "transduction" and includes reference to the incorporation of a nucleic acid sequence into a eukaryotic or prokaryotic cell where the nucleic acid sequence may be incorporated into the genome of the cell (for example, chromosome, plasmid, plastid, or mitochondrial DNA), converted into an autonomous replicon, or transiently expressed (for example, transfected mRNA).

As used herein, a "plant cell" refers to any cell derived from a plant, including cells from undifferentiated tissue (*e.g.*, callus) as well as plant seeds, pollen, progagules and embryos.

As used herein, the terms "native" and "wild-type" relative to a given plant trait or phenotype refers to the form in which that trait or phenotype is found in the same variety of plant in nature.

As used herein, the term "modified" regarding a plant trait, refers to a change in the phenotype of a transgenic plant relative to the similar non-transgenic plant. An "interesting phenotype (trait)" with reference to a transgenic plant refers to an observable or measurable phenotype demonstrated by a T1 and/or subsequent generation plant, which is not displayed by the corresponding non-transgenic (i.e., a genotypically similar plant that has been raised or assayed under similar conditions). An interesting phenotype may represent an improvement in the plant or may provide a means to produce improvements in other plants. An "improvement" is a feature that may enhance the utility of a plant species or variety by providing the plant with a unique and/or novel quality.

An "altered pathogen resistance phenotype" refers to detectable change in the response of a genetically modified plant to pathogenic infection, compared to the similar, but non-modified plant. The phenotype may be apparent in the plant itself (*e.g.*, in growth, viability or particular tissue morphology of the plant) or may be apparent in the ability of the pathogen to proliferate on and/or infect the plant. As used herein, "improved pathogen resistance" refers to increased resistance to a pathogen.

As used herein, a "mutant" polynucleotide sequence or gene differs from the corresponding wild type polynucleotide sequence or gene either in terms of sequence or expression, where the difference contributes to a modified plant phenotype or trait. Relative to a plant or plant line, the term "mutant" refers to a plant or plant line which has a modified plant phenotype or trait, where the modified phenotype or trait is associated with the modified expression of a wild type polynucleotide sequence or gene.

As used herein, the term "T1" refers to the generation of plants from the seed of T0 plants. The T1 generation is the first set of transformed plants that can be selected by application of a selection agent, *e.g.*, an antibiotic or herbicide, for which the transgenic plant contains the corresponding resistance gene. The term "T2" refers to the generation of plants by self-fertilization of the flowers of T1 plants, previously selected as being transgenic.

As used herein, the term "plant part" includes any plant organ or tissue, including, without limitation, seeds, embryos, meristematic regions, callus tissue, leaves, roots, shoots, gametophytes, sporophytes, pollen, and microspores. Plant cells can be obtained from any plant organ or tissue and cultures prepared therefrom. The class of plants which

can be used in the methods of the present invention is generally as broad as the class of higher plants amenable to transformation techniques, including both monocotyledenous and dicotyledenous plants.

As used herein, "transgenic plant" includes reference to a plant that comprises within its genome a heterologous polynucleotide. The heterologous polynucleotide can be either stably integrated into the genome, or can be extra-chromosomal. Preferably, the polynucleotide of the present invention is stably integrated into the genome such that the polynucleotide is passed on to successive generations. A plant cell, tissue, organ, or plant into which the heterologous polynucleotides have been introduced is considered "transformed", "transfected", or "transgenic". Direct and indirect progeny of transformed plants or plant cells that also contain the heterologous polynucleotide are also considered transgenic.

Identification of Plants with an Improved Pathogen Resistance Phenotype

We used an *Arabidopsis* activation tagging screen to identify the association between the gene we have designated "PPR1 (for *P. parasitica Resistant*)," predicted to encode a protein Ethylene Response Factor 1 (ERF1)- like protein, and an altered pathogen resistance phenotype, specifically, increased resistance to the fungal pathogen *P. parasitica* (downy mildew), a biotrophic pathogen controlled by the salicylic-acid (SA) dependent resistance pathway. Briefly, and as further described in the Examples, a large number of *Arabidopsis* plants were mutated with the pSKI015 vector, which comprises a T-DNA from the Ti plasmid of *Agrobacterium tumifaciens*, a viral enhancer element, and a selectable marker gene (Weigel *et al*, supra). When the T-DNA inserts into the genome of transformed plants, the enhancer element can cause up-regulation genes in the vicinity, generally within about 10 kilobase (kb) of the insertion. T1 plants were exposed to the selective agent in order to specifically recover transformed plants that expressed the selectable marker and therefore harbored T-DNA insertions. Samples of approximately 18 T2 seed were planted, grown to seedlings, and inoculated with *P. parasitica* spores. Disease symptoms on individual plants were scored based on the number of conidiophores that emerged. Accordingly, plants on which growth of conidiophores was reduced were identified as pathogen resistant.

An *Arabidopsis* line that showed increased resistance to *P. parasitica* infection was identified. The association of the PPR1 gene with the pathogen resistance phenotype was discovered by analysis of the genomic DNA sequence flanking the T-DNA insertion in the

identified line. Accordingly, PPR1 genes and/or polypeptides may be employed in the development of genetically modified plants having a modified pathogen resistance phenotype. PPR1 genes may be used in the generation of crops and/or other plant species that have improved resistance to infection by *P. parasitica* and other oomycetes and may also be useful the generation of plant with improved resistance to fungal, bacterial, and/or other pathogens. Mis-expression of PPR1 genes may thus reduce the need for fungicides and/or pesticides. The modified pathogen resistance phenotype may further enhance the overall health of the plant.

PPR1 Nucleic Acids and Polypeptides

Arabidopsis PPR1 nucleic acid (coding) sequence is provided in SEQ ID NO:1 and in Genbank entry GI 7363407, nucleotides 8077-8811. The corresponding protein sequence is provided in SEQ ID NO:2 and in GI 8844121.

As used herein, the term "PPR1 polypeptide" refers to a full-length PPR1 protein or a fragment, derivative (variant), or ortholog thereof that is "functionally active," meaning that the protein fragment, derivative, or ortholog exhibits one or more or the functional activities associated with the polypeptide of SEQ ID NO:2. In one preferred embodiment, a functionally active PPR1 polypeptide causes an altered pathogen resistance phenotype when mis-expressed in a plant. In a further preferred embodiment, mis-expression of the functionally active PPR1 polypeptide causes increased resistance to *P. parasitica* and/or other oomycetes. In another embodiment, a functionally active PPR1 polypeptide is capable of rescuing defective (including deficient) endogenous PPR1 activity when expressed in a plant or in plant cells; the rescuing polypeptide may be from the same or from a different species as that with defective activity. In another embodiment, a functionally active fragment of a full length PPR1 polypeptide (i.e., a native polypeptide having the sequence of SEQ ID NO:2 or a naturally occurring ortholog thereof) retains one of more of the biological properties associated with the full-length PPR1 polypeptide, such as signaling activity, binding activity, catalytic activity, or cellular or extra-cellular localizing activity. Some preferred PPR1 polypeptides display DNA binding activity. A PPR1 fragment preferably comprises a PPR1 domain, such as a C- or N-terminal or catalytic domain, among others, and preferably comprises at least 10, preferably at least 20, more preferably at least 25, and most preferably at least 50 contiguous amino acids of a PPR1 protein. Functional domains can be identified using the PFAM program (Bateman A et al., 1999 Nucleic Acids Res 27:260-262; website at

pfam.wustl.edu). A preferred PPR1 fragment comprises an AP2 domain (PF00847). In SEQ ID NO:2, the AP2 domain is located at approximately amino acids 79-144 Functionally active variants of full-length PPR1 polypeptides or fragments thereof include polypeptides with amino acid insertions, deletions, or substitutions that retain one of more of the biological properties associated with the full-length PPR1 polypeptide. In some cases, variants are generated that change the post-translational processing of a PPR1 polypeptide. For instance, variants may have altered protein transport or protein localization characteristics or altered protein half-life compared to the native polypeptide.

As used herein, the term "PPR1 nucleic acid" encompasses nucleic acids with the sequence provided in or complementary to the sequence provided in SEQ ID NO:1, as well as functionally active fragments, derivatives, or orthologs thereof. A PPR1 nucleic acid of this invention may be DNA, derived from genomic DNA or cDNA, or RNA.

In one embodiment, a functionally active PPR1 nucleic acid encodes or is complementary to a nucleic acid that encodes a functionally active PPR1 polypeptide. Included within this definition is genomic DNA that serves as a template for a primary RNA transcript (i.e., an mRNA precursor) that requires processing, such as splicing, before encoding the functionally active PPR1 polypeptide. A PPR1 nucleic acid can include other non-coding sequences, which may or may not be transcribed; such sequences include 5' and 3' UTRs, polyadenylation signals and regulatory sequences that control gene expression, among others, as are known in the art. Some polypeptides require processing events, such as proteolytic cleavage, covalent modification, etc., in order to become fully active. Accordingly, functionally active nucleic acids may encode the mature or the pre-processed PPR1 polypeptide, or an intermediate form. A PPR1 polynucleotide can also include heterologous coding sequences, for example, sequences that encode a marker included to facilitate the purification of the fused polypeptide, or a transformation marker.

In another embodiment, a functionally active PPR1 nucleic acid is capable of being used in the generation of loss-of-function pathogen resistance phenotypes, for instance, via antisense suppression, co-suppression, etc.

In one preferred embodiment, a PPR1 nucleic acid used in the methods of this invention comprises a nucleic acid sequence that encodes or is complementary to a sequence that encodes a PPR1 polypeptide having at least 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95% or more sequence identity to the polypeptide sequence presented in SEQ ID NO:2.

In another embodiment a PPR1 polypeptide of the invention comprises a polypeptide sequence with at least 50% or 60% identity to the PPR1 polypeptide sequence of SEQ ID NO:2, and may have at least 70%, 80%, 85%, 90% or 95% or more sequence identity to the PPR1 polypeptide sequence of SEQ ID NO:2. In another embodiment, a PPR1 polypeptide comprises a polypeptide sequence with at least 50%, 60%, 70%, 80%, 85%, 90% or 95% or more sequence identity to a functionally active fragment of the polypeptide presented in SEQ ID NO:2, such as an AP2 domain. In yet another embodiment, a PPR1 polypeptide comprises a polypeptide sequence with at least 50%, 60 %, 70%, 80%, or 90% identity to the polypeptide sequence of SEQ ID NO:2 over its entire length and comprises an AP2 domain.

In another aspect, a PPR1 polynucleotide sequence is at least 50% to 60% identical over its entire length to the PPR1 nucleic acid sequence presented as SEQ ID NO:1, or nucleic acid sequences that are complementary to such a PPR1 sequence, and may comprise at least 70%, 80%, 85%, 90% or 95% or more sequence identity to the PPR1 sequence presented as SEQ ID NO:1 or a functionally active fragment thereof, or complementary sequences.

As used herein, "percent (%) sequence identity" with respect to a specified subject sequence, or a specified portion thereof, is defined as the percentage of nucleotides or amino acids in the candidate derivative sequence identical with the nucleotides or amino acids in the subject sequence (or specified portion thereof), after aligning the sequences and introducing gaps, if necessary to achieve the maximum percent sequence identity, as generated by the program WU-BLAST-2.0a19 (Altschul *et al.*, J. Mol. Biol. (1990) 215:403-410; website at blast.wustl.edu/blast/README.html) with search parameters set to default values. The HSP S and HSP S2 parameters are dynamic values and are established by the program itself depending upon the composition of the particular sequence and composition of the particular database against which the sequence of interest is being searched. A "% identity value" is determined by the number of matching identical nucleotides or amino acids divided by the sequence length for which the percent identity is being reported. "Percent (%) amino acid sequence similarity" is determined by doing the same calculation as for determining % amino acid sequence identity, but including conservative amino acid substitutions in addition to identical amino acids in the computation. A conservative amino acid substitution is one in which an amino acid is substituted for another amino acid having similar properties such that the folding or activity of the protein is not significantly affected. Aromatic amino acids that can be

substituted for each other are phenylalanine, tryptophan, and tyrosine; interchangeable hydrophobic amino acids are leucine, isoleucine, methionine, and valine; interchangeable polar amino acids are glutamine and asparagine; interchangeable basic amino acids are arginine, lysine and histidine; interchangeable acidic amino acids are aspartic acid and glutamic acid; and interchangeable small amino acids are alanine, serine, threonine, cysteine and glycine.

Derivative nucleic acid molecules of the subject nucleic acid molecules include sequences that hybridize to the nucleic acid sequence of SEQ ID NO:1. The stringency of hybridization can be controlled by temperature, ionic strength, pH, and the presence of denaturing agents such as formamide during hybridization and washing. Conditions routinely used are well known (see, *e.g.*, Current Protocol in Molecular Biology, Vol. 1, Chap. 2.10, John Wiley & Sons, Publishers (1994); Sambrook *et al.*, *supra*). In some embodiments, a nucleic acid molecule of the invention is capable of hybridizing to a nucleic acid molecule containing the nucleotide sequence of SEQ ID NO:1 under stringent hybridization conditions that comprise: prehybridization of filters containing nucleic acid for 8 hours to overnight at 65° C in a solution comprising 6X single strength citrate (SSC) (1X SSC is 0.15 M NaCl, 0.015 M Na citrate; pH 7.0), 5X Denhardt's solution, 0.05% sodium pyrophosphate and 100 μg/ml herring sperm DNA; hybridization for 18-20 hours at 65° C in a solution containing 6X SSC, 1X Denhardt's solution, 100 μg/ml yeast tRNA and 0.05% sodium pyrophosphate; and washing of filters at 65° C for 1 h in a solution containing 0.2X SSC and 0.1% SDS (sodium dodecyl sulfate). In other embodiments, moderately stringent hybridization conditions are used that comprise: pretreatment of filters containing nucleic acid for 6 h at 40° C in a solution containing 35% formamide, 5X SSC, 50 mM Tris-HCl (pH 7.5), 5 mM EDTA, 0.1% PVP, 0.1% Ficoll, 1% BSA, and 500 μg/ml denatured salmon sperm DNA; hybridization for 18-20 h at 40° C in a solution containing 35% formamide, 5X SSC, 50 mM Tris-HCl (pH 7.5), 5 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.2% BSA, 100 μg/ml salmon sperm DNA, and 10% (wt/vol) dextran sulfate; followed by washing twice for 1 hour at 55° C in a solution containing 2X SSC and 0.1% SDS. Alternatively, low stringency conditions can be used that comprise: incubation for 8 hours to overnight at 37° C in a solution comprising 20% formamide, 5 x SSC, 50 mM sodium phosphate (pH 7.6), 5X Denhardt's solution, 10% dextran sulfate, and 20 μg/ml denatured sheared salmon sperm DNA; hybridization in the same buffer for 18 to 20 hours; and washing of filters in 1 x SSC at about 37° C for 1 hour.

As a result of the degeneracy of the genetic code, a number of polynucleotide sequences encoding a PPR1 polypeptide can be produced. For example, codons may be selected to increase the rate at which expression of the polypeptide occurs in a particular host species, in accordance with the optimum codon usage dictated by the particular host organism (see, e.g., Nakamura Y *et al*, Nucleic Acids Res (1999) 27:292). Such sequence variants may be used in the methods of this invention.

The methods of the invention may use orthologs of the *Arabidopsis* PPR1. Methods of identifying the orthologs in other plant species are known in the art. Normally, orthologs in different species retain the same function, due to presence of one or more protein motifs and/or 3-dimensional structures. In evolution, when a gene duplication event follows speciation, a single gene in one species, such as *Arabidopsis*, may correspond to multiple genes (paralogs) in another. As used herein, the term "orthologs" encompasses paralogs. When sequence data is available for a particular plant species, orthologs are generally identified by sequence homology analysis, such as BLAST analysis, usually using protein bait sequences. Sequences are assigned as a potential ortholog if the best hit sequence from the forward BLAST result retrieves the original query sequence in the reverse BLAST (Huynen MA and Bork P, Proc Natl Acad Sci (1998) 95:5849-5856; Huynen MA *et al.*, Genome Research (2000) 10:1204-1210). Programs for multiple sequence alignment, such as CLUSTAL (Thompson JD et al, 1994, Nucleic Acids Res 22:4673-4680) may be used to highlight conserved regions and/or residues of orthologous proteins and to generate phylogenetic trees. In a phylogenetic tree representing multiple homologous sequences from diverse species (e.g., retrieved through BLAST analysis), orthologous sequences from two species generally appear closest on the tree with respect to all other sequences from these two species. Structural threading or other analysis of protein folding (e.g., using software by ProCeryon, Biosciences, Salzburg, Austria) may also identify potential orthologs. Nucleic acid hybridization methods may also be used to find orthologous genes and are preferred when sequence data are not available. Degenerate PCR and screening of cDNA or genomic DNA libraries are common methods for finding related gene sequences and are well known in the art (see, e.g., Sambrook, *supra*; Dieffenbach and Dveksler (Eds.) PCR Primer: A Laboratory Manual, Cold Spring Harbor Laboratory Press, NY, 1989). For instance, methods for generating a cDNA library from the plant species of interest and probing the library with partially homologous gene probes are described in Sambrook *et al.* A highly conserved portion of the *Arabidopsis* PPR1 coding sequence may be used as a probe. PPR1 ortholog

nucleic acids may hybridize to the nucleic acid of SEQ ID NO:1 under high, moderate, or low stringency conditions. After amplification or isolation of a segment of a putative ortholog, that segment may be cloned and sequenced by standard techniques and utilized as a probe to isolate a complete cDNA or genomic clone. Alternatively, it is possible to initiate an EST project to generate a database of sequence information for the plant species of interest. In another approach, antibodies that specifically bind known PPR1 polypeptides are used for ortholog isolation. Western blot analysis can determine that a PPR1 ortholog (i.e., an orthologous protein) is present in a crude extract of a particular plant species. When reactivity is observed, the sequence encoding the candidate ortholog may be isolated by screening expression libraries representing the particular plant species. Expression libraries can be constructed in a variety of commercially available vectors, including lambda gt11, as described in Sambrook, *et al.*, *supra*. Once the candidate ortholog(s) are identified by any of these means, candidate orthologous sequence are used as bait (the "query") for the reverse BLAST against sequences from *Arabidopsis* or other species in which PPR1 nucleic acid and/or polypeptide sequences have been identified.

PPR1 nucleic acids and polypeptides may be obtained using any available method. For instance, techniques for isolating cDNA or genomic DNA sequences of interest by screening DNA libraries or by using polymerase chain reaction (PCR), as previously described, are well known in the art. Alternatively, nucleic acid sequence may be synthesized. Any known method, such as site directed mutagenesis (Kunkel TA *et al.*, Methods Enzymol. (1991) 204:125-39), may be used to introduce desired changes into a cloned nucleic acid.

In general, the methods of the invention involve incorporating the desired form of the PPR1 nucleic acid into a plant expression vector for transformation of in plant cells, and the PPR1 polypeptide is expressed in the host plant.

An isolated PPR1 nucleic acid molecule is other than in the form or setting in which it is found in nature and is identified and separated from least one contaminant nucleic acid molecule with which it is ordinarily associated in the natural source of the PPR1 nucleic acid. However, an isolated PPR1 nucleic acid molecule includes PPR1 nucleic acid molecules contained in cells that ordinarily express PPR1 where, for example, the nucleic acid molecule is in a chromosomal location different from that of natural cells.

Generation of Genetically Modified Plants with a Pathogen Resistance Phenotype

PPR1 nucleic acids and polypeptides may be used in the generation of genetically modified plants having a modified pathogen resistance phenotype; in general, improved resistance phenotypes are of interest. Pathogenic infection may affect seeds, fruits, blossoms, foliage, stems, tubers, roots, etc. Accordingly, resistance may be observed in any part of the plant. In a preferred embodiment, altered expression of the PPR1 gene in a plant is used to generate plants with increased resistance to *P. parasitica*. In a further preferred embodiment, plants that mis-express PPR1 may also display altered resistance to other pathogens. Other oomycete pathogens of interest include *Pythium spp, Phytophthora spp, Bremia lactucae, Peronosclerospora spp., Pseudoperonospora. Sclerophthora macrospora, Sclerospora graminicola, Plasmopara viticola, and Albugo candidia.* Fungal pathogens of interest include *Alternaria brassicicola, Botrytis cinerea, Erysiphe cichoracearum, Fusarium oxysporum, Plasmodiophora brassica, Rhizoctonia solani, Colletotrichum coccode, Sclerotinia spp., Aspergillus spp., Penicillium spp., Ustilago spp.*, and *Tilletia spp.* Bacterial pathogens of interest include *Agrobacterium tumefaciens, Erwinia tracheiphila, Erwinia stewartii, Xanthomonas phaseoli , Erwinia amylovora, Erwinia carotovora, Pseudomonas syringae, Pelargonium spp, Pseudomonas cichorii, Xanthomonas fragariae, Pseudomonas morsprunorum, Xanthomonas campestris.*

The methods described herein are generally applicable to all plants. Although activation tagging and gene identification is carried out in *Arabidopsis*, the PPR1 gene (or an ortholog, variant or fragment thereof) may be expressed in any type of plant. In preferred embodiments, the invention is directed to crops including maize, soybean, cotton, rice, wheat, barley, tomato, canola, turfgrass, and flax. Other crops include alfalfa, tobacco, and other forage crops. The invention may also be directed to fruit- and vegetable-bearing plants, plants used in the cut flower industry, grain-producing plants, oil-producing plants, and nut-producing plants, among others.

The skilled artisan will recognize that a wide variety of transformation techniques exist in the art, and new techniques are continually becoming available. Any technique that is suitable for the target host plant can be employed within the scope of the present invention. For example, the constructs can be introduced in a variety of forms including, but not limited to as a strand of DNA, in a plasmid, or in an artificial chromosome. The introduction of the constructs into the target plant cells can be accomplished by a variety of techniques, including, but not limited to *Agrobacterium*-mediated transformation, electroporation, microinjection, microprojectile bombardment calcium-phosphate-DNA

co-precipitation or liposome-mediated transformation of a heterologous nucleic acid. The transformation of the plant is preferably permanent, *i.e.* by integration of the introduced expression constructs into the host plant genome, so that the introduced constructs are passed onto successive plant generations. Depending upon the intended use, a heterologous nucleic acid construct comprising a PPR1 polynucleotide may encode the entire protein or a biologically active portion thereof.

In one embodiment, binary Ti-based vector systems may be used to transfer polynucleotides. Standard *Agrobacterium* binary vectors are known to those of skill in the art, and many are commercially available (e.g., pBI121 Clontech Laboratories, Palo Alto, CA).

The optimal procedure for transformation of plants with *Agrobacterium* vectors will vary with the type of plant being transformed. Exemplary methods for *Agrobacterium*-mediated transformation include transformation of explants of hypocotyl, shoot tip, stem or leaf tissue, derived from sterile seedlings and/or plantlets. Such transformed plants may be reproduced sexually, or by cell or tissue culture. *Agrobacterium* transformation has been previously described for a large number of different types of plants and methods for such transformation may be found in the scientific literature.

Expression (including transcription and translation) of PPR1 may be regulated with respect to the level of expression, the tissue type(s) where expression takes place and/or developmental stage of expression. A number of heterologous regulatory sequences (e.g., promoters and enhancers) are available for controlling the expression of a PPR1 nucleic acid. These include constitutive, inducible and regulatable promoters, as well as promoters and enhancers that control expression in a tissue- or temporal-specific manner. Exemplary constitutive promoters include the raspberry E4 promoter (U.S. Patent Nos. 5,783,393 and 5,783,394), the 35S CaMV (Jones JD *et al*, Transgenic Res (1992) 1:285-297), the CsVMV promoter (Verdaguer B *et al.*, Plant Mol Biol (1998) 37:1055-1067) and the melon actin promoter (published PCT application WO0056863). Exemplary tissue-specific promoters include the tomato E4 and E8 promoters (U.S. Patent No. 5,859,330) and the tomato 2AII gene promoter (Van Haaren MJJ *et al.*, Plant Mol Bio (1993) 21:625-640). In one preferred embodiment, PPR1 expression is under the control of a pathogen-inducible promoter (Rushton *et al.*, The Plant Cell (2002) 14:749-762)

In one preferred embodiment, PPR1 expression is under control of regulatory sequences from genes whose expression is associated with the CsVMV promoter.

In yet another aspect, in some cases it may be desirable to inhibit the expression of endogenous PPR1 in a host cell. Exemplary methods for practicing this aspect of the invention include, but are not limited to antisense suppression (Smith, *et al.*, Nature (1988) 334:724-726; van der Krol et al., Biotechniques (1988) 6:958-976); co-suppression (Napoli, *et al.*, Plant Cell (1990) 2:279-289); ribozymes (PCT Publication WO 97/10328); and combinations of sense and antisense (Waterhouse, *et al.*, Proc. Natl. Acad. Sci. USA (1998) 95:13959-13964). Methods for the suppression of endogenous sequences in a host cell typically employ the transcription or transcription and translation of at least a portion of the sequence to be suppressed. Such sequences may be homologous to coding as well as non-coding regions of the endogenous sequence. Antisense inhibition may use the entire cDNA sequence (Sheehy *et al.*, Proc. Natl. Acad. Sci. USA (1988) 85:8805-8809), a partial cDNA sequence including fragments of 5' coding sequence, (Cannon et al., Plant Molec. Biol. (1990) 15:39-47), or 3' non-coding sequences (Ch'ng et al., Proc. Natl. Acad. Sci. USA (1989) 86:10006-10010). Cosuppression techniques may use the entire cDNA sequence (Napoli et al., *supra*; van der Krol et al., The Plant Cell (1990) 2:291-299), or a partial cDNA sequence (Smith et al., Mol. Gen. Genetics (1990) 224:477-481).

Standard molecular and genetic tests may be performed to further analyze the association between a gene and an observed phenotype. Exemplary techniques are described below.

1. DNA/RNA analysis

The stage- and tissue-specific gene expression patterns in mutant versus wild-type lines may be determined, for instance, by in situ hybridization. Analysis of the methylation status of the gene, especially flanking regulatory regions, may be performed. Other suitable techniques include overexpression, ectopic expression, expression in other plant species and gene knock-out (reverse genetics, targeted knock-out, viral induced gene silencing [VIGS, see Baulcombe D, (1999) Arch Virol Suppl 15:189-201]).

In a preferred application expression profiling, generally by microarray analysis, is used to simultaneously measure differences or induced changes in the expression of many different genes. Techniques for microarray analysis are well known in the art (Schena M *et al.*, Science (1995) 270:467-470; Baldwin D *et al.*, Cur Opin Plant Biol. (1999) 2(2):96-103; Dangond F, Physiol Genomics (2000) 2:53-58; van Hal NL *et al.*, J Biotechnol (2000) 78:271-280; Richmond T and Somerville S, Curr Opin Plant Biol (2000) 3:108-116). Expression profiling of individual tagged lines may be performed. Such analysis can identify other genes that are coordinately regulated as a consequence of the

overexpression of the gene of interest, which may help to place an unknown gene in a particular pathway.

2. Gene Product Analysis

Analysis of gene products may include recombinant protein expression, antisera production, immunolocalization, biochemical assays for catalytic or other activity, analysis of phosphorylation status, and analysis of interaction with other proteins via yeast two-hybrid assays.

3. Pathway Analysis

Pathway analysis may include placing a gene or gene product within a particular biochemical, metabolic or signaling pathway based on its mis-expression phenotype or by sequence homology with related genes. Alternatively, analysis may comprise genetic crosses with wild-type lines and other mutant lines (creating double mutants) to order the gene in a pathway, or determining the effect of a mutation on expression of downstream "reporter" genes in a pathway.

Generation of Mutated Plants with a Pathogen Resistance Phenotype

The invention further provides a method of identifying plants that have mutations in endogenous PPR1 that confer increased pathogen resistance, and generating pathogen-resistant progeny of these plants that are not genetically modified. In one method, called "TILLING" (for targeting induced local lesions in genomes), mutations are induced in the seed of a plant of interest, for example, using EMS treatment. The resulting plants are grown and self-fertilized, and the progeny are used to prepare DNA samples. PPR1-specific PCR are used to identify whether a mutated plant has a PPR1 mutation. Plants having PPR1 mutations may then be tested for pathogen resistance, or alternatively, plants may be tested for pathogen resistance, and then PPR1-specific PCR is used to determine whether a plant having increased pathogen resistance has a mutated PPR1 gene. TILLING can identify mutations that may alter the expression of specific genes or the activity of proteins encoded by these genes (see Colbert et al (2001) Plant Physiol 126:480-484; McCallum et al (2000) Nature Biotechnology 18:455-457).

In another method, a candidate gene/Quantitative Trait Locus (QTLs) approach can be used in a marker-assisted breeding program to identify alleles of or mutations in the PPR1 gene or orthologs of PPR1 that may confer increased resistance to pathogens (see Foolad et al., Theor Appl Genet. (2002) 104(6-7):945-958; Rothan *et al.*, Theor Appl Genet (2002) 105(1):145-159); Dekkers and Hospital, Nat Rev Genet. (2002) Jan;3(1):22-

32). Thus, in a further aspect of the invention, a PPR1 nucleic acid is used to identify whether a plant having increased pathogen resistance has a mutation in endogenous PPR1 or has a particular allele that causes the increased pathogen resistance.

While the invention has been described with reference to specific methods and embodiments, it will be appreciated that various modifications and changes may be made without departing from the invention. All publications cited herein are expressly incorporated herein by reference for the purpose of describing and disclosing compositions and methodologies that might be used in connection with the invention. All cited patents, patent applications, and sequence information in referenced websites and public databases are also incorporated by reference.

EXAMPLES

EXAMPLE 1

Generation of Plants with a Pathogen Resistance Phenotype by Transformation with an Activation Tagging Construct

Mutants were generated using the activation tagging "ACTTAG" vector, pSKI015 (GI 6537289; Weigel D *et al.*, *supra*). Standard methods were used for the generation of *Arabidopsis* transgenic plants, and were essentially as described in published application PCT WO0183697. Briefly, T0 *Arabidopsis* (Col-0) plants were transformed with *Agrobacterium* carrying the pSKI015 vector, which comprises T-DNA derived from the *Agrobacterium* Ti plasmid, an herbicide resistance selectable marker gene, and the 4X CaMV 35S enhancer element. Transgenic plants were selected at the T1 generation based on herbicide resistance. T2 seed was collected from T1 plants and stored in an indexed collection, and a portion of the T2 seed was accessed for the screen.

Approximately 18 T2 seeds from each of the greater than 40,00 lines tested were planted in soil. The seed were stratified for three days and then grown in the greenhouse for seven days. The seedlings were inoculated with approximately $1x10^5$ conidia per ml *P. parasitica* spores and incubated in a dew room at 18°C and 100% humidity for 24 hours. The plants were then moved to a growth room at 20°C and 60% relative humidity with ten-hour long light period for six days. Individual plants were evaluated for the presence or absence of conidiophores on cotyledons. Lines in which at least a single plant showed no conidiophore growth were re-tested in a secondary screen by releasing three sets of 18 seed and screening for resistance to *P. parasitica* growth as before.

Lines in which a significant number of plants showed no conidiophores after infection were subjected to a tertiary screen. Approximately 54 T2 seed were released, planted individually and infected with *P. parasitica* as before. The plants were evaluated for the number of conidiophores growing on a single cotyledon and ranked by the following scoring system: a score of 0 indicates 0 conidiophores per cotyledon, 1 indicates 1-5 conidiophores per cotyledon, 2 indicates 6-10 conidiophores per cotyledon, 3 indicates 11-20 conidiophores per cotyledon, and 4 indicates greater than 20 conidiophores per cotyledon

The ACTTAG line designated W000030248 was identified as having an increased resistance phenotype. Specifically, 8.8% of individual plants showed no conidiophores in the secondary screen. In the tertiary screen, 7 plants scored as 0, 14 scored 1, 20 scored 2, 9 scored 3 and 2 scored 4. Control (wild-type Col-0) plants displayed significantly greater susceptibility; 0 plants scored 0, 1 plant scored 1, 1 plant scored 2, 6 plants scored 3, and 9 plants scored 4.

Plants from line W000030248 also displayed altered morphological phenotypes. In the T1 generation, these plants displayed leaf petiole and leaf epidermis phenotypes. In the T2 generation, these plants displayed leaf petiole and leaf epidermis phenotypes, as well as late flowering and reduced size.

The insertion mutation was predicted to have a dominant or semi-dominant effect. Gentoyping of individual W000030248 T2plants analyzed in the tertiary screen indicated that plants that were homozygous for the insert were more resistant to *P.parasitica* infection than were heterozygotes, which were more resistance than wild-type plants.

The dominant *P.parasitica* resistance phenotype in W000030248 is heritable. Approximately 54 individual W000030248 plants from two T3 families were analyzed for resistance to *P. parasitica*. The results indicated that plants homozygous or heterozygous for the insert showed comparable resistance to infection by *P.parasitica*.

EXAMPLE 2

Characterization of the T-DNA Insertion in Plants Exhibiting the Altered Pathogen Resistance Phenotype.

We performed standard molecular analyses, essentially as described in patent application PCT WO0183697, to determine the site of the T-DNA insertion associated with the increased pathogen resistance phenotype. Briefly, genomic DNA was extracted from plants exhibiting increased pathogen resistance. PCR, using primers specific to the

pSKI015 vector, confirmed the presence of the 35S enhancer in plants from line W000030248, and Southern blot analysis verified the genomic integration of the ACTTAG T-DNA and showed the presence of a single T-DNA insertion in the transgenic line.

Plasmid rescue was used to recover genomic DNA flanking the T-DNA insertion, which was then subjected to sequence analysis.

The sequence flanking the left T-DNA border was subjected to a basic BLASTN search and/or a search of the *Arabidopsis* Information Resource (TAIR) database (available at the arabidopsis.org website), which revealed sequence identity to BAC clone F9P14 (GI 7363407), mapped to chromosome 1. The T-DNA inserted at nucleotide 2971 of F9P14. Sequence analysis revealed that the T-DNA had inserted in the vicinity (*i.e.*, within about 10 kb) of the gene whose nucleotide sequence is presented as SEQ ID NO: 1 and GI 7363407, nucleotides 8077-8811, and which we designated PPR1. Specifically, the T-DNA inserted approximately 5 kb 5' to the coding sequence of the PPR1 gene.

EXAMPLE 3

Analysis of *Arabidopsis* PPR1 Sequence

The amino acid sequence predicted from the PPR1 nucleic acid sequence is presented in SEQ ID NO:2 and GI 8844121. PFAM analysis identified an AP2 DNA binding domain located at amino acids 79-144. A serine-rich region is located near the carboxy terminus.

Sequence analyses were performed with BLAST (Altschul *et al.*, *supra*) and PFAM (Bateman *et al.*, *supra*), among others. BLAST analysis indicated that SEQ ID NO:2 has similarity to the DNA binding protein S25-XP1 from *Nicotiana tabacum* (GI 7489116 and GI 1732406), the *Arabidopsis* Ethylene Response Factor (ERF1; GI 4128210, GI 4128208, and GI 15229405), a putative DNA binding protein from *Oryza sativa* (GI 19034045), an ethylene response factor ERF1-like protein from *Oryza sativa* (GI 24060083), and transcription factor TSRF1 from *Lycopersicon esculentum* (GI 23452024). The top BLAST hit was GI 22326027 (At2g31230; SEQ ID NO:3), which is annotated as "ethylene response factor, putative", and shares 70% overall identity with PPR1 (SEQ ID NO:2). GI 22326027, like PPR1, has an AP2 domain and has a serine-rich region near the carboxy terminus. ERF1 lacks a similar serine-rich region. GI22326027 also shares high identity with PPR1 at the carboxy terminal 32 amino acids (approx. 73% identity). By comparison, ERF1 shares only about 33% identity with PPR1 in this region.

Thus, orthologs of PPR1 may be expected to contain serine-rich regions and share greater sequence identity with PPR1 (SEQ ID NO:2) compared to ERF1 (GI 4128210).

EXAMPLE 4

Confirmation of Phenotype/Genotype Association

RT-PCR analysis showed that the PPR1 gene was overexpressed in pathogen-resistant plants from line W000030248. Specifically, RNA was extracted from tissues derived from plants exhibiting the pathogen resistance phenotype and from wild type COL-0 plants. RT-PCR was performed using primers specific to the sequence presented as SEQ ID NO:1, to other predicted genes in the vicinity of the T-DNA insertion, and to a constitutively expressed actin gene (positive control). The results showed that plants displaying the pathogen resistance phenotype over-expressed the mRNA for the PPR1 gene, indicating the enhanced expression of the PPR1 gene is correlated with the pathogen resistance phenotype.

EXAMPLE 5

Recapitulation of Pathogen Resistance Phenotype

Arabidopsis plants of the Ws ecotype were transformed by agrobacterium mediated transformation with a construct containing the coding sequences of the PPR1 gene (At1g06160, gi|15221402) behind the CsVMV promoter and in front of the nos terminator or a control gene unrelated to pathogen resistance. Both of these constructs contain the *npt*II gene to confer kanamycin resistance in plants. T1 seed was harvested from the transformed plants and transformants selected by germinating seed on agar medium containing kanamycin. Kanamycin resistant transformants were transplanted to soil after 7 days and grown for 4 weeks. Control plants were germinated on agar medium without kanamycin, transplanted to soil after 7 days and grown in soil for 4 weeks

To evaluate pathogen resistance, transformants and control plants were sprayed with a suspension of 1 x 10^5 conidia per ml of *P. parasitica*, incubated at 100% humidity for 1 day, and grown for 6 more days in the growth room After this growth period, plants were rated for severity of disease symptoms. A score of 0 means the leaves had 0-10% of the number of conidiophores growing on the leaf surface as a fully susceptible plant, 1 means 10-25% the number of conidiophores, 2 means 25-50%, 3 means 50-75% and 4 means 75-100%. Fifty-two plants transformed with PPR1, 50 plants transformed with the control gene and 10 control plants were examined.

Degree-of-infection scores were obtained from each plant tested. As a group, the PPR1 transformants were more resistant to *P. parasitica* infection than control plants. In PPR1 transformants, 11.5% were scored as 0, 15.4% as 1, 21.5% as 2, 30.8% as 3 and 21.5% as 4. In plants transformed with the control gene, only 4% scored as 0, 4% as 1, 6% as 2, 4% as 3, while 82% scored as 4. In control plants, 0% scored 0, 1, and 2, 10% scored 3 and 90% scored 4. These data show that plants over-expressing PPR1 are significantly more resistant to *P. parasitica* infection than wild-type plants.

Further analysis of plants constitutively expressing PPR1, showed that they constitutively express endogenous PDF1.2, a pathogenesis related (PR) protein that is a molecular marker for the jasmonic acid (JA)-dependent resistance pathway. The JA-dependent resistance pathway controls necrotrophic fungi and oomycetes such as *Alernaria brassicola* or *Botrytis cinerea.* These plants also constitutively express endogenous PR1, a marker for the salicylic acid (SA)-dependent resistance pathway. The SA-dependent resistance pathway controls bacterial pathogens such as *Pseudomonas spp. Xanthomonas spp.,* and *Erwinia* and biotrophic fungi and oomycetes such as *Erysiphe cichoracearum* and *Peronospora parasitica.* Thus, plants genetically modified to overexpress a PPR1 ortholog, may be similarly expected to also overexpress endogenous PR1 and PDF1.2 relative to non-transgenic plants and be resistant to pathogens that are controlled by both the SA- and JA-dependent resistance pathways. In contrast, plants that constitutively express ERF1 constitutively express PDF1.2 but not PR1 and are resistant to pathogens controlled by only the JA-dependent resistance pathway (Berrocal-Lobo *et al.*, The Plant Journal (2002) 29(1):23-32).